DOCUMENT RESUME

ED 358 123 TM 019 898

AUTHOR Oldham, Ben R.
TITLE High Stakes Assessment: A Local District Perspective.
PUB DATE Apr 93
NOTE 28p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Accountability; Decision Making; *Educational Assessment; Educational Change; Educational Legislation; Elementary Secondary Education; Program Implementation; Rewards; *School Districts; Staff Development; *State Legislation; Student Evaluation; *Test Use
IDENTIFIERS *High Stakes Tests; Kentucky Education Reform Act 1990; Monitoring; Performance Based Evaluation; Reform Efforts

ABSTRACT

The Kentucky Education Reform Act legislated by the 1990 General Assembly created a high-stakes school performance accountability system to monitor the progress of implementation. One major component of the accountability system is a schedule of consequences designed to reward those schools making sufficient progress in improving student performance and to sanction schools that maintained current achievement levels or declined. The cognitive and non-cognitive components of the assessment system are described, and the impact is discussed from a local district perspective. The following system components are highlighted: (1) the use of assessment results to make individual decisions about students; (2) the scoring rule applied to student performance; (3) the impact of performance events; (4) teacher workload; (5) differential student achievement growth; and (6) the influence on staff development. Only time will tell if the mandates of the Kentucky Education Reform Act will produce a better educated product of the public schools. Initial activities tend to focus on a quick fix, the greatest impact on performance in the most efficient manner. It is likely that long-term professional development activities will be those that are characterized by what truly makes a difference in the classroom. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

BEN R. OLDHAM

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED358123

HIGH STAKES ASSESSMENT:

A LOCAL DISTRICT PERSPECTIVE

Ben R. Oldham

Division of Research and Assessment
Fayette County Public Schools
Lexington, Kentucky 40502
Phone: 606/281-0125

Paper presented at the annual meeting of the
American Educational Research Association
Atlanta, GA, April, 1993

TM019898

Abstract

The Kentucky Education Reform Act legislated by the 1990 General Assembly created a high-stakes school performance accountability system to monitor the progress of implementation. One major component of the accountability system is a schedule of consequences designed to reward those schools making sufficient progress in improving student performance and to sanction the schools that maintained current achievement levels or declined.

The purpose of this paper is to describe the cognitive and noncognitive components of the assessment system and discuss the impact from a local school district perspective. The uses of assessment results to make individual decisions about students, the scoring rule applied to student performance, the impact of performance events, the teacher workload, differential school achievement growth, and the influence on staff development are highlighted.

HIGH STAKES ASSESSMENT: A LOCAL DISTRICT PERSPECTIVE

Background

In Kentucky it started with a lawsuit. In 1985 the superintendents of 66 of Kentucky's 177 school districts with the lowest per pupil property values filed suit against the Commonwealth of Kentucky, charging that the public school system was "inadequate and inequitable". The landmark decision was rendered in favor of the plaintiffs. Kentucky's public school system was ruled unconstitutional in that it failed to comply with the state constitutional mandate that, "The General Assembly shall by appropriate legislation, provide for an efficient system of common schools throughout the state". The result of that decision was an order to the Kentucky General Assembly to fund education at a higher level and to develop a new system of public schools to meet constitutional requirements (Luttrell, 1990).

Funding for the public schools and other education and humanities programs was increased from $1.63 billion to $2.02 billion for 1990-91, a 22% increase (Luttrell, 1990). As a rule, greater accountability follows an increase in funding. The resulting Kentucky Education

Reform Act (KERA) mandated changes in a number of areas. The impact of the change is to be evaluated by the greatest change of all, the assessment of the performance of Kentucky's public school students.

Description

The Kentucky assessment program includes cognitive and noncognitive measures. Those measures, described below, are summarized into a score called the Accountability Index. The Accountability Index for a biennium is compared to the Accountability Index from the prior biennium to determine the success a school has made toward achieving seventy-five "valued outcomes". All valued outcomes were not assessed in the initial assessment cycle. The program will be incrementally increased each biennium until the full implementation in 1995-1996.

Cognitive Index

The cognitive index, determined by assessing all students in grades 4, 8, and 12, contributes five-sixths of the weight to the Accountability Index. It is calculated by combining three assessment types - a

transitional test, performance events, and portfolio scores (currently writing only but will be expanded to include student mathematics products for the 1993-94 assessment cycle).

Transitional tests are similar in design to those administered as part of the National Assessment of Educational Progress (NAEP). There are five subtests: writing, reading, mathematics, science, and social studies. For grades 8 and 12, time allocations are 90 minutes per subtest with a permitted extension of 45 minutes for those students who have not completed a subtest. The fourth grade subtests have a 60 minute suggested completion time with a 30 minute extension for those students who have not completed the test. For the 1991-92 transitional assessment, the reading, mathematics, science, and social studies subtests contained 55 multiple-choice and 4 open-response items. While the writing subtest had the same time constraints, the students responded to a writing prompt rather than a multitude of test items. Students were asked to select one from two randomly assigned topics. Prewriting activities were encouraged but the final draft was the only writing scored.

Performance events were administered to small groups of students in mathematics, science, and social studies. The assessment required one class period where an outside assessor administered the performance tasks. In grade 4 all students were assessed, while in grades 8 and 12 a sample of students was required to respond to a performance event. Beginning with the 1992-93 school year all students will respond to a sampled performance task. There were twelve performance events administered, with students randomly assigned to one of the tasks. In most cases, small groups of students discussed a situation or problem in mathematics, science, or social studies and offered group solutions for approximately 20 minutes. At the conclusion of the group period, students broke a seal on an answer folder and responded individually to the problem.

Portfolios are samples of best student writing (to be expanded to products in mathematics in 1992-93). Students are offered a great deal of flexibility in the portfolio entries but a wide representation of student writing is required. Entries such as a personal narrative, a written reaction to a cultural or sports event, a writing piece that predicts an outcome, defends

a position, solves a problem, draws a conclusion or creates a model, a short story, poem, play or other piece of original fiction, and a letter to the reviewer discussing the writer's reflecting on the pieces in the portfolio are examples of portfolio entries. In all there must be seven entries. Teachers grade the portfolios from their classes and rescore a random sample of portfolio entries from other teachers' classes to provide a mechanism for monitoring the reliability.

Noncognitive Index

The contribution of noncognitive indicators makes up one sixth of the Accountability Index. The noncognitive index is derived by combining attendance rates (all grades), retention rates (all grades), dropouts (middle and high school only), transition (graduates only), and reduction of barriers to learning.

Attendance rates are calculated by dividing the aggregate days absence for a student population by the aggregate days membership for the school year.

Retention rates are calculated by dividing the number of students retained by the student membership.

Dropouts are calculated by dividing the number of

students who withdraw from school identified by standard withdrawal codes in addition to the students who do not return to school in the fall by the number of students who were in membership. Dropout calculations are only for schools that serve students in grade 7 or above.

Transition refers to a successful move of graduates from high school to a postsecondary experience. A successful postsecondary experience is defined as graduates attending college or vocational/technical school, students gainfully employed, students who have joined the military, and students who are homemakers. The high school is responsible for confirming the postsecondary status of graduates. The number identified as having made a successful transition is divided by the number of graduates to determine the transition rate.

"Removal of barriers to learning" refers to situations that keep students from achieving at the highest levels. The barriers may be physical or emotional. At this point the barriers factor is not included in the calculation of the Accountability Index.

Accountability Index

The cognitive and noncognitive factors are combined

into a number called the Accountability Index. A Threshold or improvement goal for the next biennium is established for each school from the Accountability Index by determining the gap between the current Accountability Index and an Accountability Index of 100 at the end of a twenty year period. That gap is divided by ten to determine how much Accountability Index growth is required to keep a school on target. Schools will be rewarded monetarily for exceeding the biennial goal by at least one percent. Sanctions will be imposed on those schools that fail to meet the Threshold. The sanctions increase in severity as the Accountability Index declines. (see 1991-92 Technical Report)

Impact on Local School Districts

A number of factors influence the usefulness of the assessment results. Since the assessment program is high-stakes by design, significant attention will be devoted to the improvement of results regardless of the impact on real achievement. School quality will be defined by responses to assessment tasks. Following is a discussion of the major factors and the resulting impact of the Kentucky assessment program from a local school district

perspective.

Impact on Individual students

The greatest benefit to students from the Kentucky Education Reform Act could occur as a result of the change in the way teachers must now establish expectations for students. Citing Kentucky School Law (KRS 158.6455, 1992) "It is the intent of the General Assembly that schools succeed with all students and receive appropriate consequences to that success". Schools can no longer use socioeconomic status, ethnicity, or home environment as excuses why students cannot achieve. This statement of law supports the outcome-based philosophy that departs from the traditional bell-curve thinking. Educators must change the way of thinking about all students' potential to achieve in an outcome-based model as described by Spady (1992). Grading must be based on what students know and are able to do. Textbooks must be replaced by identified valued outcomes. Curriculum tracking must be reduced. The materials and instructional methods used in programs for gifted students must be accessible to all students.

Individual student results are reported for both

portfolios and the transitional tests (the tests taken in booklets in a more traditional format). The transitional test is made up of multiple-choice and open-response items. To gain a broader sampling of the curriculum, matrix-sampled items (those items that are unique to a particular test form) were included reading, mathematics, science, and social studies. The matrix-sampled items were used for the calculation of school results but only common items were included in the determination of individual performance status.

Individual students were classified in one of four groupings - novice, apprentice, proficient, or distinguished according to their performance on three common open-response items per subtest. While matrix sampling provides a sufficient item pool for acceptable reliability at the school building level, reliability is not adequate to permit decisions about individual students (1991-92 Technical Report, 1992).

For schools this presents several problems. Because of the amount of time and budget dollars devoted to the performance assessment program, it is unreasonable to devote additional assessment time and the resulting fiscal impact to two major testing programs. Schools at

all levels use standardized assessment information as one piece in a decision making equation. Academic program placement decisions are made as students proceed from elementary school to middle school and middle school to high school. Teacher recommendations are helpful but a reliable standardized measure is invaluable in the decision making process. An assessment system that provides data with acceptable reliability only at the group level limits the usefulness for the decision makers in the schools.

Additionally, most school districts have developed programs with achievement criteria required for admission. Programs for the academically gifted and talented, Chapter I, Duke Talent Search, etc. are examples of the programs that historically have required standardized norm-referenced tests. The elimination of standardized tests for student selection for these programs may be desirable but if test criteria are removed something must be used to fill the requirement.

Parents have become accustomed to receiving test scores presented in a normative format. While assessment scores are presented and interpreted in a performance format, parents continue to ask, "yes, but how does my

child stack up nationally" questions. The charge is sometimes leveled that schools are hiding something. The parent education component of performance assessment is monumental. There is no major objection to performance assessment but parents do not seem ready to give up normative comparisons.

Kentucky schools are especially concerned about a high-stakes assessment program that could result in sanctions being imposed on a school while the students are not held accountable for their performance. Results, unreliable at the individual level, cannot be used as an incentive to motivate students to expend their greatest efforts. The results mean everything to the school but nothing to the students. This concern is especially evident at the senior high school level. High school students are involved in high-stakes school assessment the second semester of the senior year. Real student performance changes could occur and go undetected in a high-stakes for schools but a no-stakes environment for students.

Some schools in Kentucky have established a performance requirement for students. That performance requirement requires students to submit a portfolio in

order to graduate from high school. In time, the districts will require a specified level of performance. Such a procedure will place some of the onus on students and will hopefully make them more effective participants in the high-stakes assessment process.

Impact on Instruction

Kentucky public school children have been required to take norm-referenced achievement tests in designated grades dating back to the Educational Improvement Act of 1978. The focus until the 1991-92 school year has been on improving results derived from multiple-choice tests. Using skill-based item analysis reports, schools identified areas of concern and addressed those concerns in classroom activities. The emphasis, however, was on developing test-taking strategies to improve multiple-choice test performance.

With the assessment component of the Kentucky Education Reform Act being performance-based, the preparation activities differ markedly. If the assessment of performance represents what students should know and be able to do, then the classroom activities will ultimately reflect the authentic assessment program.

Because the KERA assessment program is based on improvement regardless of the achievement status, teachers must alter what goes on in the classroom to meet the biennial improvement goal. The resulting alteration in instructional practice will not occur without a substantial professional development component.

To estimate a school's performance in an authentic setting, students in grades 4, 8, and 12 are brought into a room, usually the school library, where they respond to one of twelve performance tasks. The students are randomly assigned to the performance tasks. Students work in small groups for approximately twenty minutes. At that time the assessor instructs the students to work on an individual response to the task. Students only respond to one of the twelve tasks. With four tasks per subject area in mathematics, science, and social studies measuring the performance of a school with each student being given the opportunity to encounter one task, the small number of students per task will limit the reliability regardless of the quality of the performance tasks and the interrater consistency. An external validity issue also comes into question. Can the results of four performance tasks be generalized to represent achievement in

mathematics, science, or social studies? Schools observed large performance differences between mathematics, science, and social studies achievement that resulted from the random assignment of students to the tasks.

Kentucky educators expected differences in assessment practice with a dominance of performance assessment activities. The lack of measurement reliability listed in the Technical Report (1992) has limited the performance assessment component of the Accountability Index to approximately ten percent. The impact on instruction is limited, therefore, to the amount the various components of the assessment program contribute to the Accountability Index.

Communication of knowledge in reading, mathematics, science, and social studies is a primary goal in the Kentucky Education Reform Act. There is concern among teachers that, while assessing the communication of knowledge is important, the direct assessment of knowledge is not being given sufficient consideration. Items that assess writing in response to reading, mathematics, science, and social studies comprise approximately 57 percent of the Accountability Index on which school will be evaluated. With the inclusion of

portfolio scores the writing requirement of the Accountability Index is approximately 74 percent of the total score. While teachers generally support the importance of written communication of knowledge, knowledge of the subject in and of itself seems to be inadequately represented in the model. Teachers in mathematics and science are particularly concerned about the allocation of instructional time to writing.

Assessment Across Subject Domains

To the credit of the Kentucky Department of Education and the company that was awarded the assessment contract, Advanced Systems in Measurement & Evaluation, Inc. (ASME), local school district personnel have been heavily involved in the development and review of test items, performance tasks, and the establishment of a scoring standard.

The decision rule to classify students as novice, apprentice, proficient, or distinguished was developed by subject area specialists from Kentucky schools under the direction of the professionals from ASME. While the process was good, the subject area specialists developed the scoring rules for each subtest independently. When

the scoring rule is applied to student performance, the resulting distribution of students across the achievement categories for different subject areas is not linked. The problem arises at the school level upon the receipt of results. The school does not know whether the achievement distribution differences between reading, mathematics, science, and social studies are a result of real differences in academic performance or are a result of a higher or lower standard being applied to the student performance. It would be feasible for professionals developing the scoring rule for a subtest area to establish more challenging standards to get additional attention devoted to that subject area.

Teacher Involvement and Workload

Teachers and other school-based professionals are generally supportive of a program that assesses what students know and are able to do. There is a concern, however, that many of the required activities do not directly support improved instructional practice.
As an example, schools must verify the post graduation status of former students. That verification can be from various sources, the most time consuming of which

involves contacting the student or parents by telephone.

Another time consuming responsibility requires secondary schools to verify the status of students who have withdrawn from a school. The withdrawal or transfer status of all students must be verified. In many cases this verification requires only a request for student records from another school. However, in cases where a student does not return to school in the fall and a receiving school does not request records, the school must spent time tracking the enrollment status of former students. No one denies the importance of locating all students and placing them in programs that lead to a high school diploma. The problem is that additional responsibilities are placed on schools without commensurate increases in personnel to perform those tasks.

An additional to the workload for teachers involves the multiple grading of portfolios in writing and mathematics in grades 4, 8 and 12. Teachers are beginning to understand the philosophy that portfolios must become a part of the classroom assessment process. Teachers have been evaluating student products for centuries. The difference in this evaluation process is the interrater

reliability factor. That is, all teachers must be assigning the same or nearly the same score when they rate portfolios. To check and improve the rating consistency, a sample of portfolios from each teacher must be rescored. It is the rescoring that bothers teachers most. Philosophically teachers understand, but rescoring takes additional valuable time. If the scores are discrepant beyond a point defined by ASME, the teacher must rescore all portfolios in the class. Besides the embarrassment of being singled out, the process involves teacher work beyond what was required in prior years without release time provided.

In grade 4 teachers must maintain and evaluate multiple portfolios for students - one in writing and one for mathematics. Those teachers in the elementary grades have the rescoring problem compounded with two portfolios to manage for each student.

Growth in Achievement

The Kentucky assessment program is developed around the premise that all students can learn and achieve at a high level. The assessment design, therefore, establishes a common achievement goal for all schools at the end of

a twenty year period. Since the starting achievement point for each school is different but the ultimate goal is the same for all, schools must realize different achievement gains as they progress. A high achieving school is required to make smaller annual achievement gains to reach the goal than the low achieving schools. The schools serving the most disadvantaged communities have the greatest challenge to overcome the barriers to learning.

While many educators do not disagree with the this philosophy, those who serve more difficult student populations are expected to exceed the educational growth of the advantaged populations without adequate support.

Influence on Staff Development

Measurement drives instruction. That is, the kinds of things measured and the methodology used to measure will influence what is taught and how it is taught (Popham, 1987). This is magnified in a high-stakes assessment environment. If measurement is skill-based and assessment items are constructed in a multiple-choice format, then instructional strategies will be developed to prepare students to represent themselves well. The

instructional focus to prepare students to take a multiple-choice test will likely be on skills, with the classroom assessment dominated by multiple-choice tests. If this represents good educational practice, then traditional assessment practice will suffice.

The KERA program was designed to assess in such a manner to encourage improved classroom practice. The scores from the transitional tests, those tests administered in booklets come from written responses from students. Students must communicate what they know and are able to do. Being able to indicate what they know is no longer adequate.

Preparing teachers for the high-stakes assessment program by providing staff development activities that are directly related to the assessment program could be one of the most beneficial outcomes of the reform act. Teachers within Kentucky have been given a listing of "valued outcomes" outlining what students should know and be able to do. Teachers have been and will be apprised of the process used in the assessment of students. The charge to local districts is to provide the staff development experiences to promote exemplary instructional practices that will be reflected in the

assessment results.

Staff development activities must, therefore, be designed to improve the communication process in all areas of the curriculum. Teachers are retrained to teach writing in response to mathematics, science, and social studies. They are taught how to elicit higher order thinking behavior. Teachers are taught to develop assessment tasks moving from using verbs like "list", "define", and "identify" to verbs like "explain", "compare and contrast", and defend". By changing classroom assessment strategies to prepare for the high-stakes assessment, improved instructional practice should result.

The performance activities included in the KERA assessment package require students to work as a group to react to or solve an authentic problem. It is not in the teachers' or schools' best interests to utilize a lecture presentation of material if the high-stakes assessment requires a substantially different mode of addressing a problem. Teachers, to position their students for optimal performance, must alter practice to make it consistent with assessment. That will necessitate extensive retraining for most classroom teachers with an emphasis

on cooperative learning and assessing the student products.

The training of teachers in a high-stakes portfolio assessment environment must focus on developing consistent assessment practice, that is, high interrater reliability. This training pulls teachers from different school sites to a common location for instruction and dialogue. There is a tremendous advantage to be gained in getting teachers together to discuss what constitutes acceptable, exemplary and unacceptable performance. The dialogue should have immediate impact on instructional practice. From the classroom teachers' perspective, interrater reliability is a minor consideration but the dialogue between classroom professionals directed toward evaluating best student products represents what the measurement community had hoped would be a teacher outcome. The discussions that inevitably result on strategies to alter classroom practice to ensure the achievement of certain desired student outcomes is a primary benefit.

The difficulties districts are experiencing in the Commonwealth of Kentucky are associated with a large number of legislated mandates in the Kentucky Educational

Reform Act and an inadequate amount of time devoted to staff training. In a high-stakes assessment environment, educators seem to be focusing initially on what can be done to have the greatest impact on student outcomes in the most efficient manner, that is, a quick fix. Those kinds of "high scoring" strategies will likely only last for one biennium in Kentucky. The Kentucky reform plan is based on biennial improvement regardless of the level achieved. So while the initial activities are directed toward Accountability Index improvement activities, the long-term professional development programs will likely be characterized by what truly makes a difference in the classroom.

Conclusion

Only time will tell if the mandates of the Kentucky Educational Reform Act will produce a better educated product of the public schools in Kentucky. Any system is initially more painlessly implemented if it is done in a "bottom-up" rather than a "top-down" manner, but one of the realities of high-stakes assessment programs is that it is mandated. Classroom practice has changed little over the decades regardless of the quality of the

research. Research-into-practice, while a key element for our profession is not occurring in any large degree to the alter instructional methodology in the classroom. The staff development and staff retraining that is commonplace in other professions must be a characteristic in education. The research is being carried out, but there is a decided gap in what has been shown to work and what is being implemented. Possibly a high-stakes assessment like the one legislated in Kentucky can have an influence on the professional practice of users of research.

References

Kentucky Department of Education. (1992). Kentucky Instructional Results Information System: 1991-92 Technical Report. Frankfort, KY: Kentucky Department of Education.

Kentucky Department of Education. (1992). Kentucky School Laws 1992. Frankfort, KY: Kentucky Department of Education.

Luttrell, W. (1990). Education Reform In Kentucky: The Funding. School Governance & Administration Exchange, 9, 1-3.

Popham, W.J. (1987). The Merits of Measurement-Driven Instruction. Phi Delta Kappan, 68, 679-682.

Spady, W.G. (1992). It's Time to Take a Close Look at Outcome-Based Education. Outcomes, 11 (2), 6-13.